Synthetic Image Detection Using a Modern CNN and Noise Patterns

Pedro Henrique Ralph Arruda
Orientador: Tsang Ing Ren

[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brazil

phra@cin.ufpe.br, tir@cin.ufpe.br

Abstract. *Today, people enjoy unprecedented access to state-of-the-art artificial intelligence models. This generates great opportunities for innovation and development but also several problems. Artists working with images are now worried about deep learning models trained to replicate their styles, and public figures are increasingly concerned about AIs being able to mimic their appearance and voice. In general, people are now concerned about whether they can trust anything they see, hear, or read on the internet. One very present problem is to trust if an image is true or computer generated. In the paper, we evaluate whether new methods of image synthesis can still be recognized by deep learning models. This work evaluates the new developments in CNNs and pattern recognition applied to synthetic image detection. The results present improvements in the proposed model compared to previous works on the efficiency gains of new architectures of convolutional neural networks and on the generalization potential of models trained through pattern identification methods.*

Resumo. *Hoje, as pessoas desfrutam de acesso sem precedentes a modelos de inteligência artificial de última geração. Isso gera grandes oportunidades de inovação e desenvolvimento, mas também vários problemas. Artistas que trabalham com imagens agora estão preocupados com modelos de aprendizado profundo treinados para replicar seus estilos, e figuras públicas estão cada vez mais preocupadas com a capacidade de IAs de imitar sua aparência e voz. Em geral, as pessoas agora estão preocupadas se podem confiar em qualquer coisa que vejam, ouçam ou leiam na internet. Um problema muito presente é confiar se uma imagem é verdadeira ou gerada por computador. No artigo, avaliamos se novos métodos de síntese de imagens ainda podem ser reconhecidos por modelos de aprendizado profundo. Este trabalho avalia os novos desenvolvimentos em CNNs e reconhecimento de padrões aplicados à detecção de imagens sintéticas. Os resultados apresentam melhorias no modelo proposto em relação a trabalhos anteriores sobre os ganhos de eficiência de novas arquiteturas de redes neurais convolucionais e sobre o potencial de generalização de modelos treinados por meio de métodos de identificação de padrões.*

1. Introduction

Presently several different models can produce synthetic images. This task was enabled mainly due to the advent of generative models. The history of generative models began with Generative Adversarial Networks (GAN) in 2014[Goodfellow et al. 2020]

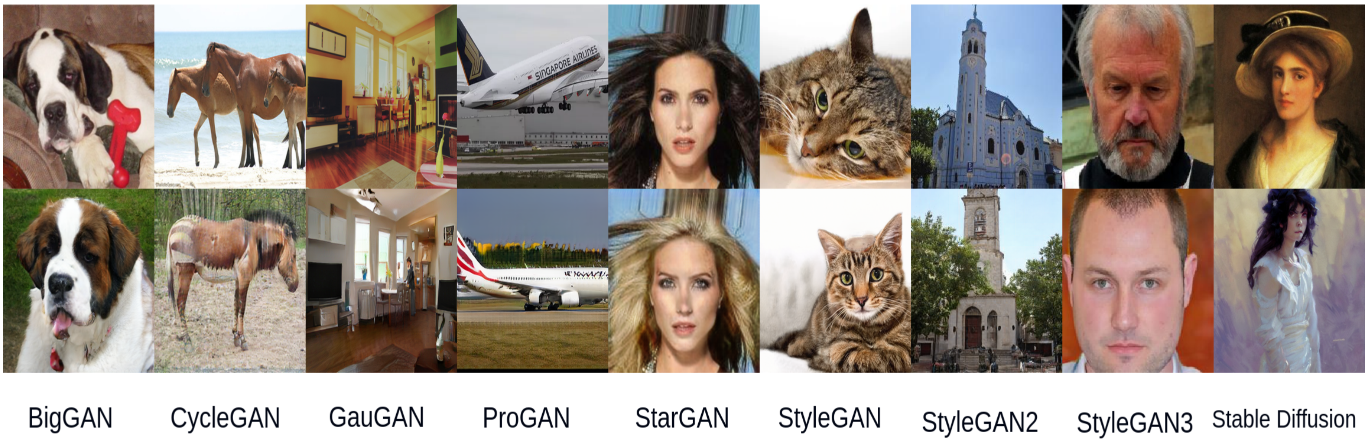

Figure 1. Examples of images evaluated in this study were obtained from various generative methods. The upper row of the images comprises real images, while the bottom row displays synthetic images. These synthetic images were generated using selected generative models.

and together with Style Transfer published in 2016[Gatys et al. 2016] progressed into StyleGAN[Karras et al. 2019]. Furthermore, this procedure evolved to text-to-image models that started with another development, the CLIP model in 2021 [Radford et al. 2021], that tries to describe by the text an image received as input through multimodal learning.

The first text-to-image models used CLIP together with a GAN network to iteratively approximate the image generated by the GAN network to the text given as input ranked by CLIP. The greatest success of this era of machine learning generators was the VQGAN+CLIP developed in 2021 and peer-reviewed in 2022 [Crowson et al. 2022] that showed for the first time the real power of generation possessed by well-developed and trained neural networks. VQGAN+CLIP created the first boom of interest from normal people in an area that was previously only of academic interest, with many people discovering the best prompts to give themselves as entry to the network, and many others are publishing modifications and their versions of VQGAN+CLIP.

The next advance in the field was the development of the so-called diffusion models made public in early 2021 by OpenAI [Ramesh et al. 2021]. These models showed performance and result far superior to the GAN models. Diffusion models were used to create several commercial programs and websites and continue to be in active development. Currently, they can create images very close to reality or works of art rivaling to those used for training the neural network.

Moreover, when generative models became good enough to raise concerns that they might one-day fool humans, interest in detecting and classifying the images created by these models arose. These detection methods are divided into two approaches: artifact detection and data-based detection. However, even with good results in recognizing images generated by the same generative models that created the training images, the detection methods have difficulty recognizing images generated by other models.

Data-based approaches to detection are very dependent on the availability of large datasets and models trained on a large number of real and fake images to learn common features of the generative models that created the fake images. A purely data-driven approach quickly encounters the generalization problem since images generated by unseen models will not be detected reliably. Resizing and lossy compression methods of images found on the internet make data augmentation a common practice on models using this approach. Examples include:[Wang et al. 2020], [Gragnaniello et al. 2021], [Rahman et al. 2023]. All of them use a large and diverse dataset and extensively use data augmentation to deal with the generalization problem.

Artifact detection approaches try to use artifacts left by the processes of the generative network to identify images. Detection based on spatial domain artifacts quickly becomes ineffective as the quality of the images generated by a method improves. With the current rate of development of synthetic images, the detection methods can be effective for a short time. Frequency domain artifacts are interesting since each type of generative method leaves a unique fingerprint that can be used to identify it. Artifact detection approaches are still vulnerable to the generalization problem. They are even more affected by resizing and lossy compression since the traces left by the generative network are often very small and fragile. Furthermore, complications arise because some synthesis methods like StyleGAN3 [Karras et al. 2021] work to eliminate even the normally invisible frequency artifacts. Examples can be found in reference [Durall et al. 2020] and [Frank et al. 2020]. B. Liu et al. [Liu et al. 2022a] use frequency analysis differently, instead of learning the artifact pattern of a given generative method, they try to learn the common features of real images.

This article presents a modern CNN architecture detection model trained using images from an "old" generative model ProGAN [Karras et al. 2018] as a training base. These experiments show that synthetic image detection is still possible by using the advances in CNN models and Learned Noise Patterns (LNP) proposed by B. Liu et al. [Liu et al. 2022a]. Here we applied a very recently proposed CNN architecture, the ConvNeXt [Liu et al. 2022b].

2. Related Works

The development of new image synthesis models has led to a growing interest in detecting images produced by these models, and significant progress has already been made in this area.

J. Frank et al. [Frank et al. 2020] show that Fourier transforms reveal patterns that can be used as a form of signature of each generative model. Marra et al. [Marra et al. 2019] showed that generative models leave fingerprints on their generated images, and that pre-trained CNN classifiers perform better than CNNs trained only for detecting generated images [Marra et al. 2018]. Wang et al. [Wang et al. 2020] showed that large and diverse datasets are fundamental for good results. Data Augmentation techniques such as blurring and simulating lossy compression are essential to guarantee results in real use cases. Yu et al. [Yu et al. 2019] showed that GANs have unique signatures, and they can be used to distinguish them. B. Liu et al. [Liu et al. 2022a] found common patterns between real images instead of using the patterns created by generative networks.

The detection of synthetic images is facilitated by identifying underlying patterns that distinguish them from real images. It is, therefore, imperative to preserve these patterns during the training and testing stages of the model. Resizing operations may inadvertently alter these patterns, leading to inaccurate results. Gragnaniello et al. [Gragnaniello et al. 2021] demonstrated the importance of preserving these patterns by achieving improved results through the removal of two layers of downsampling from the Resnet network, as compared to the earlier work by Wang [Wang et al. 2020].

The primary issue faced by these models is their inability to generalize effectively, which is further exacerbated by the rapid advancement of generative models. Despite the emergence of new diffusion models, the identifying patterns of a model still persist, leading to poor classification results when trained on images generated by a different model. This finding is supported by Corvi et al. [Corvi et al. 2022].

The images evaluated in this study are presented in Fig. 1. The upper row includes real images, while the bottom row features synthetic images from various datasets.

3. Methodology

3.1. Learned Noise Patterns (LNP)

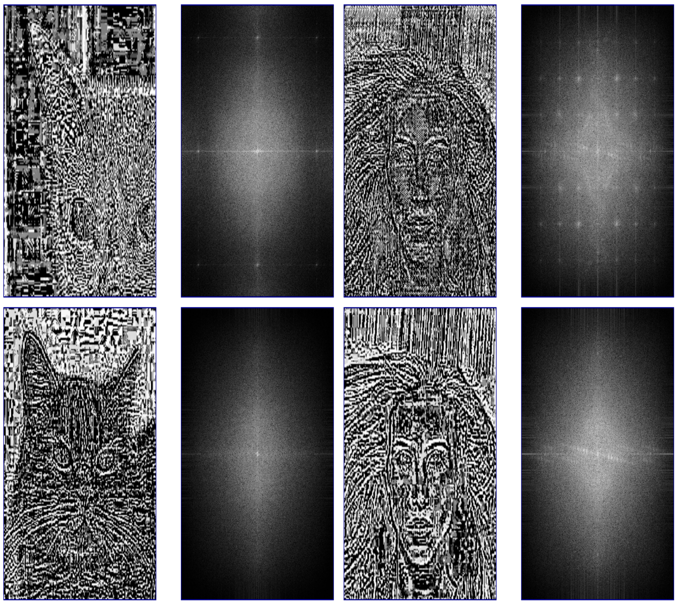

Figure 2. The upper row consists of images that had their LNP extracted and their Fourier transforms, which consist of generated images by the StyleGAN3 and StarGAN models, respectively. The bottom row consists of real images. Unlike StyleGAN3, which generates an entirely new image, StarGAN enhances an existing image so a more direct comparison is possible. A grid-like pattern is clearly visible on the Fourier transform of the generated images.

Proposed by B. Liu et al. [Liu et al. 2022a], the LNP are patterns created by the different incidences of light during the creation of the image by a camera. These patterns do not have a periodicity in a real image, but in synthetic images, this periodicity is present due to the processes necessary for creating the image. The revealed pattern has a grid-like shape that is more visible in some images than others but is very noticeable in the Fourier space. LNPs can be used in two ways to extract as much information as possible from an image, **LNP Amplitude Spectrum** is given by the two-dimensional Fourier transform of the image, thus making image frequency analysis possible. We can observe that real images have a very similar frequency domain, which makes it possible to distinguish the generated images that have their distinct patterns. Fig.2 shows examples of the LNP features and its Fourier transforms for synthetic and real images.

The **LNP Phase Spectrum** seeks to extract structural information from the image. Through experiments, better results are found using the phase spectrum of the LNP instead of the phase spectrum of the original image.

3.2. Extracting LNPs

The extraction of LNPs is done through a denoising neural network, which is especially effective for our application since they manage to extract the noise from an image without affecting its content. Among the various noise removal networks, better results were obtained with CycleISP [Zamir et al. 2020], which transforms RGB images to RAW and then back to RGB to have more realistic noise removal results. LNPs are then obtained from the difference between the original and denoised images.

3.3. ConvNeXt

ConvNeXt showed excellent results in image classification, object detection, and semantic segmentation applications [Liu et al. 2022b]. Its good results and the fact that it is a pure CNN allow us to compare it with the ResNet network and verify whether the gains in accuracy and efficiency are noticeable even in its most minor version of the ConvNeXt, ConvNeXt-Tiny. The results obtained by Rahman et al. [Rahman et al. 2023] using a larger version of ConvNeXt, while not attributable to architecture choice alone, are encouraging.

Recalling that the LNPs, as well as the artifacts left by the generator models, can be disturbed or erased by processing operations on the images, we reduced the stride factor of the ConvNeXt stem block to 2 instead of 4 in order to reduce the loss of information from the images during convolution. Rahman et al. showed that such a modification results in a significant performance improvement.

4. Experiments and Results

4.1. Datasets

The 20 classes of the dataset provided by [Wang et al. 2020] were used for training, containing images with common objects, vehicles, animals, and persons. It has 726K images divided equally between images generated by ProGAN and real images, with other 8K images for a validation set divided equally as the training set.

A part of the dataset provided by [Wang et al. 2020] was used as a test set. It consists of 63K images generated by various types of GANs (StarGAN [Choi et al. 2018], CycleGAN [Zhu et al. 2017], GauGAN [Park et al. 2019], BigGAN [Brock et al. 2019], ProGAN[Karras et al. 2018], StyleGAN [Karras et al. 2019], StyleGAN2 [Karras et al. 2020]) divided equally between generated and real images. Two additional datasets were also tested with more recent methods. The first consists of 20K images generated by Stable Diffusion selected from DiffusionDB [Wang et al. 2022] and 20K images of real works of art selected from ArtBench-10 [Liao et al. 2022]. Moreover,

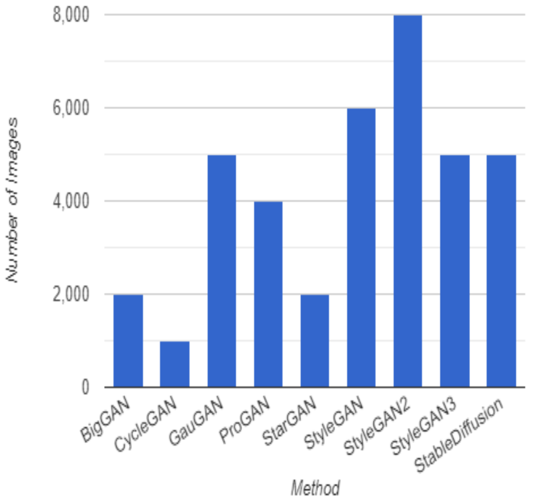

Figure 3. Number of images contained in each method on the test dataset.

a second dataset comprises 10K images generated by StyleGAN3 [Karras et al. 2021] and 10K real images, all obtained from the ArtiFact [Rahman et al. 2023] dataset were also used. Fig.1 shows some images from the selected generative methods.

In order to assess the impact of dataset size on the performance of the models, a smaller dataset was created by selecting a subset of images from the ProGAN dataset. Specifically, this reduced dataset contains only 181K images, in contrast to the original ProGAN dataset which has a much larger number of images. The generative models were then trained on this smaller dataset to determine whether the size of the dataset has an effect on their ability to produce synthetic images. Fig.3 presents the results of this experiment, showing the number of synthetic images generated by each method on the test dataset. It is clear that the number of synthetic images varies significantly across the different methods, indicating that some models may be better suited for smaller datasets than others. This information is useful for researchers and practitioners who may be working with limited data resources and need to optimize their training process accordingly.

4.2. Experiments

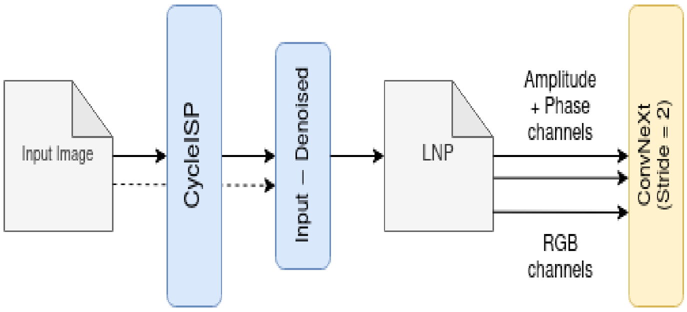

Figure 4. Proposed model structure.

The experiments were done using a ConvNeXt-Tiny [Liu et al. 2022b] network

Table 1. Comparison on the accuracy of a model trained on a reduced dataset with the accuracy of the same model tested on a stable diffusion dataset.

Methods	Accuracy
Resnet	51.2
ConvNeXt-Tiny	53.5
Resnet+LNP	67.9
ConvNeXt-Tiny+LNP (ours)	**70.8**

pre-trained on the ImageNet dataset, and all the images have dimensions of 256x256, using an Adam optimizer and an initial learning rate of 10^{-4}. In the reduced dataset, a Resnet50 [He et al. 2016] network pre-trained also on the ImageNet dataset was also used. Fig.4 illustrates the proposed model's structure with all its components.

4.3. Study on a reduced datataset

An important aspect in the evaluation of machine learning models is their ability to generalize to new datasets. To test the impact of the training dataset size on the accuracy of image detection models, we trained a model on a reduced version of the ProGAN dataset containing only 181K images. To evaluate the performance of the detection methods, an initial test was conducted on a reduced dataset. The dataset, comprising only 181K images, was used to investigate the impact of dataset size on the models' performance. Table 1 shows the results of the early testing. The ConvNeXt+LNP method demonstrated the highest accuracy of 70.8%, outperforming the original standard ResNet method by 20.5%. These results indicate that the ConvNeXt+LNP method is a promising approach for detecting synthetic images and can potentially be further optimized for improved performance. The use of LNP proves to improve considerably the results compared with not using this feature.

4.4. Comparisons

The models utilized in this study were trained on the complete ProGAN dataset and augmented with techniques such as a 10% probability of Gaussian blurring (sigma = 0.0∼3.0) and JPEG quality modifications (30∼100). The obtained results are presented in Tables 2, 3, and 4. It is worth noting that the models developed by Wang et al. and B. Liu et al. performed consistently with the results reported in their respective papers [Wang et al. 2020] [Liu et al. 2022a], as well as with the outcomes of similar studies conducted by other researchers [Gragnaniello et al. 2021] [Corvi et al. 2022].

The results achieved by the ConvNeXt-Tiny models are consistent with their performance improvements over ResNet in ImageNet classification tests, as reported in previous studies [Liu et al. 2022b]. Specifically, the ConvNeXt-Tiny model exhibited slightly superior performance compared to ResNet50 on most of the tests conducted in our experiments. These findings are particularly noteworthy given the challenges posed by synthetic image detection, which require models to distinguish between images with subtle differences in texture, color, and pattern.

Our proposed ConvNeXt-Tiny+LNP model shows a very similar performance gain over B. Liu et al. model, but a more in-depth discussion of some results are worth-

Table 2. The comparison of the accuracy with other state-of-the-art methods.

Methods	Big GAN	Cycle GAN	Gau GAN	Pro GAN	Star GAN	Style GAN	Style GAN2	Style GAN3	Stable Diffusion	Average
Resnet [Liu et al. 2022a]	73.2	87.7	82.9	**100**	94.7	92.5	86.4	57.2	53.2	80.8
ConvNeXt-Tiny (ours)	76.0	88.2	85.7	**100**	93.8	90.8	90.2	57.5	53.5	81.7
Resnet+LNP [Liu et al. 2022a]	89.8	93.2	82.4	99.4	99.9	89.2	91.4	48.5	65.9	84.4
ConvNeXt-Tiny+LNP (ours)	80.1	90.4	77.2	99.5	**100**	90.1	**96.9**	51.3	75.5	84.5
Voting Ensemble (ours)	79.5	88.8	85.5	**100**	95.3	92.2	87.4	57.0	53.3	82.1
MLP Ensemble (ours)	**92.3**	**94.4**	**85.5**	**100**	98.2	**98.9**	94.6	**66.6**	**82.0**	**90.3**

Table 3. The comparison of the precision with other state-of-the-art methods.

Methods	Big GAN	Cycle GAN	Gau GAN	Pro GAN	Star GAN	Style GAN	Style GAN2	Style GAN3	Stable Diffusion	Average
Resnet [Liu et al. 2022a]	87.2	95.3	91.1	**100**	99.0	**99.9**	99.4	**78.5**	71.2	91.2
ConvNeXt-Tiny (ours)	95.3	**98.3**	**97.3**	**100**	99.3	99.7	99.7	78.3	75.8	93.7
Resnet+LNP[Liu et al. 2022a]	**96.5**	98.7	89.2	99.9	99.9	99.4	99.7	44.9	84.1	90.2
ConvNeXt-Tiny+LNP (ours)	89.2	97.1	85.0	**100**	**100**	96.2	**99.9**	64.5	**90.6**	**94.1**
Voting Ensemble (ours)	78.5	86.8	83.6	**100**	94.8	92.2	87.4	45.8	53.3	88.3
MLP Ensemble (ours)	88.5	93.9	85.9	**100**	98.8	98.8	94.6	57.3	78.3	90.1

Table 4. The comparison of the F1-score with other state-of-the-art methods.

Methods	Big GAN	Cycle GAN	Gau GAN	Pro GAN	Star GAN	Style GAN	Style GAN2	Style GAN3	Stable Diffusion	Average
Resnet [Liu et al. 2022a]	0.742	0.869	0.832	0.999	0.952	0.923	0.862	0.277	0.137	0,732
ConvNeXt-Tiny (ours)	0.766	0.883	0.861	**1.0**	0.928	0.911	0.899	0.04	0.136	0.713
Resnet+LNP [Liu et al. 2022a]	0.898	0.930	0.829	0.993	0.999	0.879	0.906	0.04	0.506	0.775
ConvNeXt-Tiny+LNP (ours)	0.802	0.902	0.795	0.995	**1.0**	0.898	**0.968**	0.110	0.694	0.796
Voting Ensemble (ours)	0.748	0.879	0.838	**1.0**	0.951	0.916	0.856	0.062	0.124	0.708
MLP Ensemble (ours)	**0.929**	**0.943**	**0.886**	**1.0**	0.987	**0.989**	0.943	**0.452**	**0.791**	**0.883**

while. BigGAN and GauGAN performance was lower than expected. Artifacts generated by these methods may be hard for the network to notice, but the excellent BigGAN results of the ResNet+LNP method contradict this conclusion. Therefore, further investigation is necessary. StyleGAN3 was a challenge to all methods, and those based on LNP detecting had worse results. This shows that the measures taken to minimize artifacts in StyleGAN3-generated images were effective.

4.5. Ensemble Models

In this study, we conducted an analysis of the images that were correctly and incorrectly detected by each of the models used, as illustrated in Figures 5 and 6, respectively. Our results showed that there were noticeable differences in the images that each model detected correctly and incorrectly. These differences suggest that it may be possible to improve overall performance by using ensemble methods that leverage the strengths of multiple models to capitalize on these performance differences. By combining the outputs of multiple models, it may be possible to achieve higher overall accuracy and improve the reliability of the detection process.

We investigated the effectiveness of ensemble modeling using two different approaches to improve the detection of synthetic images. The first approach was a simple majority vote between the prior detection methods, while the second approach was a Stacking ensemble consisting of a Multi-Layer Perceptron (MLP) trained on the outputs of the prior methods.

The results of our experiments showed that the simple majority vote ensemble method had worse performance compared to using one of the LNP-enhanced methods individually. On the other hand, the Stacking ensemble method showed excellent results and outperformed the individual methods in terms of accuracy on most tests. However, we observed that the Stacking ensemble suffered slightly in precision.

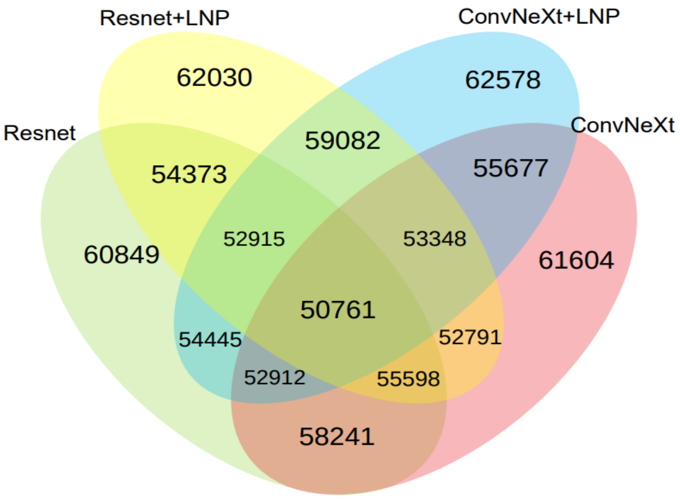

Figure 5. The Venn diagram showing the number of images correctly predicted by each method.

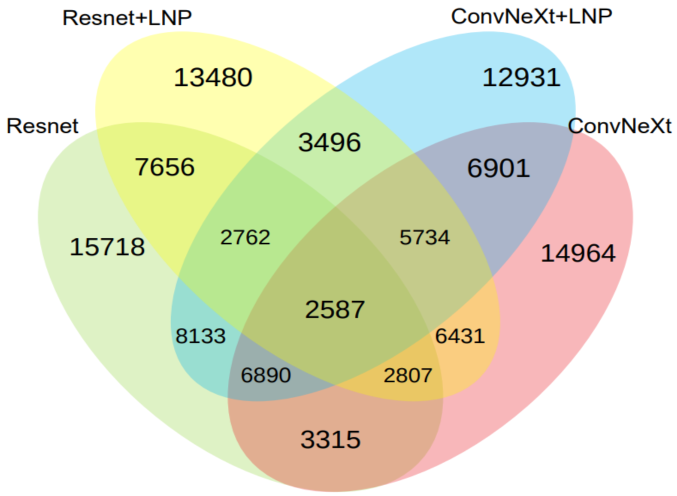

Figure 6. The Venn diagram showing the number of images incorrectly predicted by each method.

5. Conclusion and Future Work

The focus of this paper was to evaluate the effectiveness of ConvNeXt-Tiny and LNP extraction for detecting synthetic images generated by both older and newer generative models. Although there is currently no universal method available for reliably detecting synthetic images, our findings suggest that the combination of ConvNeXT-Tiny and LNP extraction shows excellent results when used in conjunction with each other.

Frequency analysis, which involves the examination of patterns in the Fourier transform of an image, has shown to be effective in detecting synthetic images. However, as newer generative models attempt to minimize artifacts such as those found in the StyleGAN3 model, the effectiveness of frequency analysis may become more challenging to perform effectively. Nonetheless, our results indicate that ConvNeXt-Tiny and LNP extraction remain promising methods for the detection of synthetic images. Further research could explore the effectiveness of these methods on a broader range of generative models and examine the potential limitations of these approaches in more detail.

The initial results obtained from testing our ConvNeXt-Tiny model with LNP extraction on the detection of generative models have shown great potential. Moreover, the findings of [Rahman et al. 2023] also indicate the effectiveness of LNP extraction in detecting synthetic images. Despite the promising results, we acknowledge that there is still a lot of room for improvement. In particular, with access to better hardware, we would like to evaluate the performance of larger versions of ConvNeXt models, potentially with deeper architectures and more parameters, trained on a more extensive and modern dataset, such as the one used in [Rahman et al. 2023].

References

Brock, A., Donahue, J., and Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., and Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., and Verdoliva, L. (2022). On the detection of synthetic images generated by diffusion models. *ArXiv:2211.00680 [cs.CV]*.

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., and Raff, E. (2022). VQGAN-CLIP: Open domain image generation and editing with natural language guidance. In Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., editors, *Computer Vision – ECCV 2022*, pages 88–105, Cham. Springer Nature Switzerland.

Durall, R., Keuper, M., and Keuper, J. (2020). Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., and Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Commun. ACM*, 63(11):139–144.

Gragnaniello, D., Cozzolino, D., Marra, F., Poggi, G., and Verdoliva, L. (2021). Are gan generated images easy to detect? a critical analysis of the state-of-the-art. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, Los Alamitos, CA, USA. IEEE Computer Society.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., and Aila, T. (2021). Alias-free generative adversarial networks. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 852–863. Curran Associates, Inc.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liao, P., Li, X., Liu, X., and Keutzer, K. (2022). The artbench dataset: Benchmarking generative models with artworks. *ArXiv:2206.11404[cs.CV]*.

Liu, B., Yang, F., Bi, X., Xiao, B., Li, W., and Gao, X. (2022a). Detecting generated images by real images. In Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., editors, *Computer Vision – ECCV 2022*, pages 95–110, Cham. Springer Nature Switzerland.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022b). A convnet for the 2020s. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976.

Marra, F., Gragnaniello, D., Cozzolino, D., and Verdoliva, L. (2018). Detection of gan-generated fake images over social networks. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 384–389.

Marra, F., Gragnaniello, D., Verdoliva, L., and Poggi, G. (2019). Do gans leave artificial fingerprints? In *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 506–511, Los Alamitos, CA, USA. IEEE Computer Society.

Park, T., Liu, M.-Y., Wang, T.-C., and Zhu, J.-Y. (2019). Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*.

Rahman, M. A., Paul, B., Sarker, N. H., Hakim, Z. I. A., and Fattah, S. A. (2023). Artifact: A large-scale dataset with artificial and factual images for generalizable and robust synthetic image detection. *arXiv:2302.11970 [cs.CV]*.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8821–8831. PMLR.

Wang, S.-Y., Wang, O., Zhang, R., Owens, A., and Efros, A. A. (2020). Cnn-generated images are surprisingly easy to spot... for now. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wang, Z. J., Montoya, E., Munechika, D., Yang, H., Hoover, B., and Chau, D. H. (2022). DiffusionDB: A large-scale prompt gallery dataset for text-to-image generative models. *arXiv:2210.14896 [cs]*.

Yu, N., Davis, L. S., and Fritz, M. (2019). Attributing fake images to gans: Learning and analyzing gan fingerprints. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M.-H., and Shao, L. (2020). Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251.